4278–4288 *Nucleic Acids Research, 2016, Vol. 44, No. 9* *Published online 19 April 2016*
doi: 10.1093/nar/gkw270

A PPR protein in the PLS subfamily stabilizes the 5′-end of processed *rpl16* mRNAs in maize chloroplasts

Kamel Hammani[1,*], Mizuki Takenaka[2], Rafael Miranda[3] and Alice Barkan[3]

[1]Centre National de la Recherche Scientifique (CNRS), Institut de Biologie Moléculaire des Plantes, 12 rue du Général Zimmer, 67084 Strasbourg, France, [2]Molekulare Botanik, Universität Ulm, 89069 Ulm, Germany and [3]Institute of Molecular Biology, University of Oregon, Eugene, OR 97403, USA

ABSTRACT

Pentatricopeptide repeat (PPR) proteins are a large family of helical-repeat proteins that bind RNA in mitochondria and chloroplasts. Precise RNA targets and functions have been assigned to only a small fraction of the >400 members of the PPR family in plants. We used the amino acid code governing the specificity of RNA binding by PPR repeats to infer candidate-binding sites for the maize protein PPR103 and its ortholog Arabidopsis EMB175. Genetic and biochemical data confirmed a predicted binding site in the chloroplast *rpl16* 5′UTR to be a site of PPR103 action. This site maps to the 5′ end of transcripts that fail to accumulate in *ppr103* mutants. A small RNA corresponding to the predicted PPR103 binding site accumulates in a PPR103-dependent fashion, as expected of PPR103's *in vivo* footprint. Recombinant PPR103 bound specifically to this sequence *in vitro*. These observations imply that PPR103 stabilizes *rpl16* mRNA by impeding 5′→3′ RNA degradation. Previously described PPR proteins with this type of function consist of canonical PPR motifs. By contrast, PPR103 is a PLS-type protein, an architecture typically associated with proteins that specify sites of RNA editing. However, PPR103 is not required to specify editing sites in chloroplasts.

INTRODUCTION

Chloroplasts and mitochondria are organelles that originated from free-living bacteria via ancient endosymbiosis events. These organelles are now semi-autonomous in the sense that they have retained a small genome from their bacterial ancestor, but the majority of the proteins required for organellar biogenesis and function are encoded in the nucleus and targeted to the organelle. Many such proteins are derived from the endosymbiont and retain their ancestral functions, but many others emerged subsequently as products of nuclear-organellar coevolution. The pentatricopeptide repeat (PPR) family (1) is a particularly large protein family that arose in this way. PPR proteins are helical repeat proteins that bind RNA and influence organellar gene expression. PPR proteins are found solely in eucaryotes, and the size of the family varies dramatically among different organisms. For example, angiosperm genomes encode >400 PPR proteins, whereas metazoans encode fewer than ten (2). PPR proteins have attracted particular attention because of their importance for organelle function, organismal development and physiology, their diverse functions in organellar RNA metabolism, and their unusual mode of RNA binding (reviewed in 3). PPR proteins are made of tandem repetitions of a variable number of PPR motifs, each of which consists of approximately 35 amino acids (1) that form two alpha helices separated by a sharp turn. Consecutive repeats stack to form a solenoid structure that binds single-stranded RNA along its surface (4,5). PPR tracts bind RNA via a modular 1 repeat-1 nucleotide recognition mode, in which the identity of the bound nucleotide is determined in part by the identity of amino acids at several specific positions in the PPR motif (6,7).

Many PPR proteins are essential for photosynthesis or respiration due to their role in promoting the expression of organellar genes required for the synthesis or function of the energy transducing machineries (reviewed in 3). In addition, PPR-encoding genes are abundant among the set of nuclear genes encoding organelle-localized proteins that are essential for seed development in Arabidopsis (8). Despite their essential roles in plant physiology and development, the molecular functions of only a small fraction of PPR proteins have been precisely characterized. Molecular functions of some PPR proteins have been inferred by close examination of photosynthesis, chloroplast transcript populations and chloroplast protein synthesis in loss-of-function mutants (e.g. (9–14)). However, embryo lethality and pleiotropic effects in many PPR mutants often complicate the assignment of functions in this way. Genome-wide

*To whom correspondence should be addressed. Tel: +33 367155281; Fax: +33 367155300; Email: kamel.hammani@ibmp-cnrs.unistra.fr

RNA-immunoprecipitation assays (RIP-chip) can identify the direct RNA ligands of PPR proteins (15–18), but this method is too laborious for the systematic assignment of RNA ligands to each member of the PPR family. A breakthrough came recently with the elucidation of an amino acid code for RNA recognition by PPR proteins (6,7,19). Although current understanding of this code is not sufficient to accurately predict binding sites of many PPR proteins, it can facilitate the computational prediction of the repertoire of RNA sequences that are likely to be bound. So far, this code has been used to identify RNAs bound by several PLS-PPR proteins that act as organellar RNA editing factors (7,19–21).

In this work, we characterized the molecular function of a maize chloroplast PPR protein, PPR103, whose Arabidopsis ortholog (EMB175/AT5G03800) is essential for embryo development (8). Disruption of *ppr103* results in albino plants that lack plastid ribosomes and that die as seedlings. PPR103 has a domain architecture that is characteristic of proteins that specify sites of RNA editing in plant organelles ("PLS-E-DYW", see below). However, we found that PPR103 is not a chloroplast RNA editing factor. To characterize its function, we used the PPR code to predict RNA binding sites for PPR103 and used these predictions to direct detailed study of specific chloroplast RNAs in *ppr103* mutants. This strategy allowed us to demonstrate that PPR103 stabilizes processed *rpl16* mRNA isoforms with a 5′ end mapping a short distance upstream of the *rpl16* gene. The position of the inferred PPR103 binding site implies that PPR103 serves as a molecular blockade to 5′→3′ degradation, analogous to functions that have been ascribed to several P-type PPR proteins in chloroplasts (reviewed in 3). We propose that this defect underlies the loss of plastid ribosomes in *ppr103* mutants, and that a conserved function is likely to account for the seed developmental defect reported for Arabidopsis *EMB175* mutants. This work expands the functional repertoire ascribed to PLS-type PPR proteins, and highlights the promise offered by computational prediction for aiding the identification of PPR binding sites and for the assignment of molecular functions to a family of essential genes in plants.

MATERIALS AND METHODS

Plant material

The *ppr103* mutants were recovered in a PCR-based reverse genetic screen of a large collection of transposon-induced non-photosynthetic maize mutants (22). The *ppr4*, *hcf7* and *atp4* mutants used as controls in some experiments were described previously (17,23,24). Plants were grown on soil for ∼9 days under 16-h light, 28°C/8-h dark, 26°C cycles. RNA and protein were extracted from the second leaf of seedlings at the three-leaf stage.

Protein analyses

Immunoblots were performed on total leaf proteins as described (25). D2 protein antibody was purchased from Agrisera manufacturer. Other antibodies were described in (26).

Bioinformatic prediction of EMB175 binding sites

To predict the potential binding sites for EMB175, we used the FIMO program in the MEME suite (http://meme-suite.org/tools/fimo), which searches sequence databases for occurrences of known motifs (27). We generated a putative nucleotide binding motif for EMB175 by using the identities of the amino acids at the 6 and 1′ position (first amino acid of the subsequent C terminal PPR motif) of each PPR motif to assign a nucleotide preference according to the weighting scheme in (19). These nucleotide preference scores were used to search EMB175 RNA binding sites against the entire chloroplast genome (NC_000932.1) using the FIMO program. The predicted binding sites were ranked by *P*-values calculated by FIMO (27).

RNA analyses

Primers used for RT-PCR, generation of probes for RNA gel blot hybridizations, and primer extension reactions are described in Supplementary Table S2.

RNA editing sites were analyzed by sequencing RT-PCR products. Three micrograms of DNA-free leaf RNA were reverse transcribed using Superscript III RT and random hexamers (Invitrogen) according to the manufacturer's instructions. RT-PCR products covering each editing site were generated with specific primers. RNA gel blot hybridizations were performed on 1, 5 or 15 μg of total leaf RNA for the detection of rRNA, mRNA or sRNA respectively, as described previously (25,28).

Primer extension assays were performed following the protocol described in (29) except that the reactions did not contain ddNTPs. For the circular RT-PCR assay, 10 μg of leaf RNA was ligated at low concentration with T4 RNA ligase, ethanol precipitated and resuspended in 10 mM Tris-HCl pH 7.5, 1 mM EDTA. Two micrograms of ligated RNA was used for reverse-transcription by SuperScript III Reverse Transcriptase (Invitrogen) using 250 ng of random primers in 20 μl reaction, according to the manufacturer's protocol. The circularized *rpl16-rpl14* junction product was amplified by PCR using k96/k100 primers. Gel-purified PCR products were A-tailed following the manufacturer's instructions (Promega) in the presence of Taq DNA polymerase (5 U) and 0.2 mM dATP in a 10 μl reaction volume before being ligated into pGEM-T and sequenced.

The sRNA sequencing data were obtained by gel-purifying RNAs between ∼15 and 40 nts from maize seedling leaf RNA, generating sequencing libraries with the NEBNext Multiplex Small RNA Library Prep Set, and sequencing on an Illumina HiSeq2000 at the University of Oregon Genomics Core Facility.

Expression of recombinant PPR103

The DNA sequence coding for the predicted mature PPR103 (i.e. lacking the transit peptide) was amplified using Phusion DNA polymerase (New England Biolabs) from maize B73 leaf DNA with primers k90/k109 containing *attB* sites for Gateway® cloning. The PCR product was subcloned into the entry vector pDONR207 (Invitrogen) and sequenced before being cloned into the destination vector pHMGWA (30) following the manufac-

turer's instructions (Invitrogen). The final construct encodes mature PPR103 fused in frame to N-terminal six-histidine and Maltose-binding protein tags. This protein, rPPR103, was expressed in Rosetta 2 (DE3) pLysS cells following induction with 1 mM IPTG and overnight incubation at 17°C under constant agitation at 220 rpm. The bacterial cells were lysed in cold buffer containing 30 mM Tris-HCl pH 7.5, 1 M NaCl, 10% glycerol, 0.05% 3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS), 5 mM β-mercaptoethanol, 0.1 mM phenylmethylsulfonyl fluoride (PMSF) and EDTA-free protease inhibitor cocktails (Roche). Soluble rPPR103 was affinity-purified on an amylose column (New England Biolabs) and the fusion protein was resolved on a Superdex S200 column (GE Healthcare Life Sciences) in lysis buffer without protease inhibitors. Only a small amount of soluble rPPR103 could be recovered. Fractions containing rPPR103 were pooled and incubated overnight at 4°C with cOmplete His-tag purification resin (Roche) to remove protein contaminants that coeluted with rPPR103. The resin was washed with buffer containing 30 mM Tris-HCl pH 7.5, 0.4 M NaCl, 10% glycerol, 0.05% CHAPS, 5 mM β-mercaptoethanol and then with the same buffer containing 5 mM imidazole. rPPR103 was eluted in wash buffer with the addition of 0.5 M imidazole. The eluted rPPR103 was transferred to a buffer containing 30 mM Tris-HCl pH 7.5, 0.4 M NaCl, 10% glycerol, 0.05% CHAPS, 5 mM β-mercaptoethanol, 0.1 mM phenylmethylsulfonyl fluoride by filtration on a Sephadex G25 column (GE Healthcare Life Sciences). The purity of the tandem affinity purified protein was visualized on SDS-PAGE and Coomassie Brilliant Blue staining. The band migrating at the expected size of rPPR103 (128.6 kDa) was gel excised and analyzed by mass spectrometry (LC-MS/MS) to confirm the identity of rPPR103. Seven grams of wet-induced bacteria pellet yielded 7 μg of soluble and virtually pure rPPR103. The recombinant protein was stored at 4°C and used within 10 days.

Gel mobility shift assays

Synthetic RNAs (Integrated DNA Technologies) were purified on a denaturing polyacrylamide gel and 5′-end–labeled with [γ-32P]–ATP and T4 polynucleotide kinase. Unincorporated radiolabeled nucleotides were removed by filtration on illustra Microspin G-25 columns (GE Healthcare) followed by phenol-chloroform extraction and ethanol precipitation. Binding reactions contained 160 mM NaCl, 30 mM Tris-HCl pH 7.5, 4 mM DTT, 0.04 mg/ml BSA, 0.5 mg/ml heparin, 10% glycerol, 0.02% CHAPS, 10 units RNaseOUT (Invitrogen) and 30 pM radiolabeled RNA. Reactions were incubated for 30 min at 25°C and resolved on 5% native polyacrylamide gels. The competition assays were carried in the same conditions, except that the unlabeled competitor RNA was preincubated with the protein for 10 min before adding the radiolabeled RNA. RNA1 and 2 are RNA oligos of similar length to that of the PPR103 footprint *rpl16* oligo and their sequence derives from fragments of Arabidopsis chloroplast tRNA Asp and Ala, respectively. Results were visualized on an FLA-7000 phosphorimager. Data quantification was performed with ImageGauge software (Fujifilm).

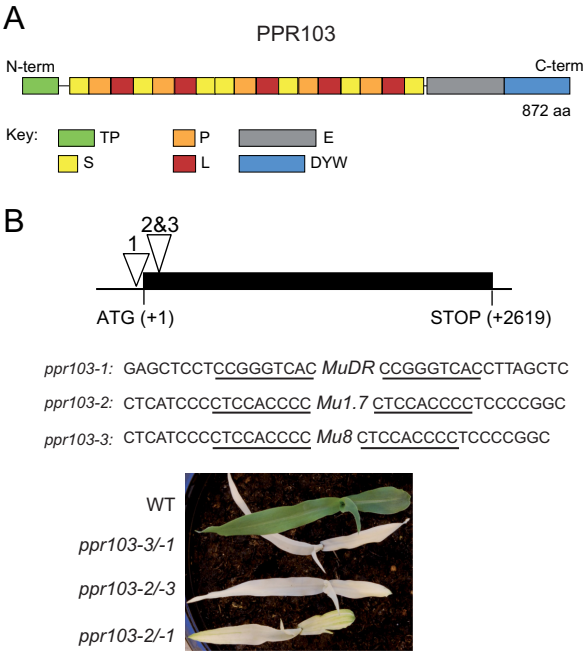


Figure 1. Overview of PPR103 protein and *ppr103* mutants. (**A**) PPR103 architecture. PPR103 is a PLS-PPR-DYW protein composed of Pure (orange), Long (red) and Short (yellow) PPR repeats with C-terminal Extended (gray) and DYW (blue) domains. Those motifs and domains are as defined in (2). The N-terminal chloroplast transit peptide (TP) is marked in green. (B) *ppr103* insertion mutants. The open reading frame lacks introns and is indicated by a black rectangle. The insertion sites are shown below, with the target site duplications underlined. The *ppr103-1/-2, -3/-1, -3/-2* plants are the heteroallelic progeny of complementation crosses. Plants were grown for ∼9 days in soil.

RESULTS

ppr103 is essential for chloroplast development in maize

PPR103 is encoded by maize gene GRMZM2G170896 and is orthologous to Arabidopsis At5g03800 (see http://cas-pogs.uoregon.edu/#/pog/11415) (31), which has been designated *EMB175* due to its essential role in embryo development (8). PPR103 has 17 PPR-like motifs that comprise a PLS tract, followed by a C-terminal E and DYW domain (Figure 1A). The majority of PLS-E-DYW proteins examined so far have been implicated in plant organellar RNA editing, a process that converts specific cytidines to uridines in organellar RNAs (32). There is increasing evidence that the C-terminal DYW domain participates in editing catalysis (33–36). Notably, the conserved eponymous DYW tripeptide is missing or mutated in PPR103 orthologs (see alignment in Supplementary Figure S1). PPR103 is predicted to localize to chloroplasts by TargetP, and was confirmed to localize to chloroplasts in a proteome study of the maize chloroplast nucleoid (37).

Three insertion alleles of *ppr103* were identified in a reverse-genetic screen of transposon-induced maize mutants in the Photosynthetic Mutant Library (22) (Figure 1B). Homozygous *ppr103-1* seedlings have an insertion

mapping 26-bp upstream of the start codon and exhibit pale yellow leaves with greening tips. The *ppr103-2* and *ppr103-3* insertions both map 35-bp downstream of the predicted start codon, but involve different members of the *Mu* transposon family; both insertions condition an albino seedling phenotype. Plants that are homozygous for any of these alleles die after the development of three to four leaves upon exhaustion of seed reserves, as is typical for non-photosynthetic maize mutants. Complementation crosses between plants heterozygous for each allele yielded ∼25% chlorophyll-deficient heteroallelic progeny (Figure 1B), confirming that the chlorophyll deficiency results from disruption of *PPR103*.

PPR103 is required for the accumulation of plastid ribosomes

The albino phenotype observed for *ppr103-2* and *-3* homozygotes is typical of maize mutants exhibiting severe plastid ribosome deficiencies. To investigate this possibility we assessed the accumulation of one core subunit of each photosynthetic enzyme complex harboring a plastid-encoded subunit (ATP synthase, photosystem II, photosystem I, cytochrome b_6f and Rubisco) in *ppr103* mutants (Figure 2A). The characterized mutants *hcf7* and *ppr5*, were included to provide a point of comparison, as they exhibit a moderate and severe loss of plastid ribosomes, respectively (18,23). The assayed proteins were undetectable in plants that were homozygous for an exon insertion (*ppr103-2* and *ppr103-3*) whereas they were reduced approximately 4-fold in plants homozygous for the 5′UTR insertion (*ppr103-1*). RNA gel blot hybridizations (Figure 2B) revealed a reduction in the levels of all chloroplast rRNAs in the progeny of *ppr103* complementation crosses, and the degree of the rRNA deficiency corresponded with the severity of the protein and pigment phenotypes (Figure 2B). These results indicate that PPR103 is required for the accumulation of plastid ribosomes.

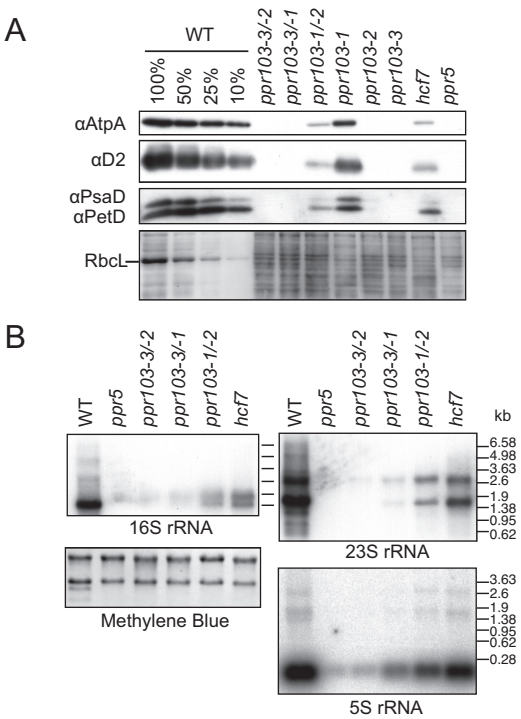


Figure 2. Evidence for loss of plastid ribosomes in *ppr103* mutants. (**A**) Immunoblot analyses of subunits of photosynthetic complexes in *ppr103* mutants. Replicate immunoblots of total leaf extract were probed with antibodies for subunits of the chloroplast ATP synthase (AtpA), photosystem II (D2), photosystem I (PsaD), and the cytochrome b_6f complex (PetD). One of the membranes was stained with Coomassie Blue (below) to demonstrate the abundance of the large subunit of Rubisco, RbcL, and to serve as a protein loading control. (**B**) Accumulation of plastid rRNAs in *ppr103* mutants. Seedling leaf RNA (1 μg) was analyzed by RNA gel blot hybridization using probes specific for the indicated plastid rRNAs. An image of one of the blots stained with methylene-blue illustrates equal loading of cytosolic rRNAs.

PPR103 is not required for RNA editing in chloroplasts

Because PPR103 is a PLS-E domain protein, we considered the possibility that it plays a role in chloroplast RNA editing. A loss of RNA editing in mRNAs that encode essential components of the chloroplast translation machinery could potentially explain the global loss of plastid translation we observed in *ppr103* mutants. To test this hypothesis, we used bulk cDNA sequencing to examine the editing status of the 27 editing sites (38,39) in the maize chloroplast transcriptome in *ppr103* mutants (Supplementary Figure S2 and Table S3). The only site that exhibited a substantial decrease in editing efficiency mapped to genome position 84 413, where editing changes ACG to AUG and creates a start codon for the *rpl2* open reading frame. However, partial editing at this site occurs in all three *ppr103* alleles (two of which are likely to be null alleles) and a similar effect on *rpl2* editing was reported for *iojap*, a maize mutant lacking plastid ribosomes (40) (Supplementary Figure S2). The reduction in editing in *iojap* mutants was proposed to be a consequence of the loss of *rpl2* splicing, which arises as a secondary effect of its defect in plastid translation (41). Thus, the reduction in *rpl2* editing in *ppr103* mutants is likely to be a pleiotropic effect resulting from their defect in plastid translation.

Computational prediction of potential PPR103 targets

Mutant phenotype is of limited use for inferring sites of action of PPR proteins that are required for the biogenesis of the plastid translation machinery because a large number of chloroplast genes contribute to plastid translation. For some proteins of this type, genome-wide RNA immunoprecipitation assays provided evidence for direct RNA binding sites (17,18). Unfortunately, our attempts to generate antibodies to PPR103 failed. As an alternative approach, we took advantage of recent advances in understanding the rules governing the RNA sequence-specificity of PPR tracts. The identities of two amino acids in each PPR motif play a major role in specifying the bound nucleotide and comprise a code for nucleotide recognition (6,7). Potential binding sites for PPR103 were predicted using a refined code that can be applied to all three types of PPR motifs (P, L and S) (19). However, PPR103 has several features that complicate this analysis. For example, some of the amino acids at positions that typically confer nucleotide specificity do not have known nucleotide binding preferences, and a 10 amino acid insertion in the ninth PPR motif complicates target prediction (Figure 3D and Supplementary Figure S1). Attempts to predict target sites based on the re-

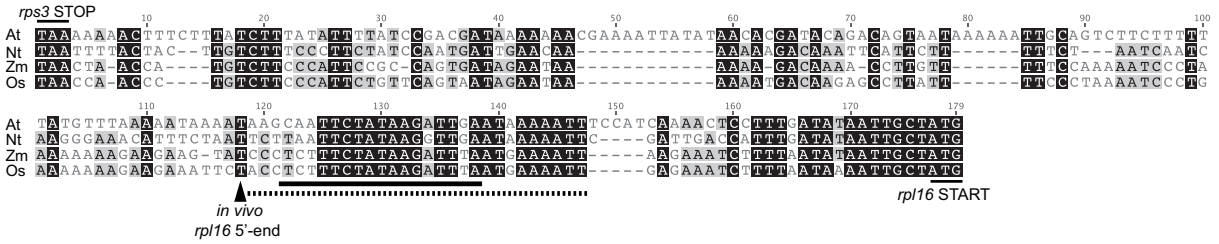
A

Motif	S	P	L	S	P	L	S	S	P	L	S	P	L	S	P	L2	S2
aa 6	N	T	V	N	N	S	N	T	N	T	T	T	T	N	N	T	T
aa 1'	T	N	S	D	D	E	D	N	T	N	S	D	D	D	D	T	E
A	0.00	0.73	0.25	0.15	0.09	0.25	0.15	0.64	0.04	0.40	0.45	0.04	0.40	0.15	0.09	0.70	0.25
C	0.78	0.05	0.25	0.23	0.25	0.25	0.23	0.06	0.69	0.19	0.19	0.00	0.19	0.23	0.25	0.16	0.25
G	0.14	0.22	0.25	0.07	0.14	0.25	0.07	0.15	0.00	0.28	0.17	0.93	0.28	0.07	0.14	0.14	0.25
U	0.09	0.00	0.25	0.55	0.52	0.25	0.55	0.16	0.27	0.14	0.18	0.03	0.14	0.55	0.52	0.00	0.25

B

		Genome position	Location	Strand	*P*-value	Sequence	S	P	L	S	P	L	S	S	P	L	S	P	L	S	P	L2	S2
	1	19121-19137	non-coding strand	+	3.06E-06	CAATTCAACGAGGTCAC	0.78	0.73	0.25	0.55	0.52	0.25	0.15	0.64	0.69	0.28	0.45	0.93	0.28	0.55	0.25	0.70	0.25
▶	2	82691-82707	*rps3-rpl16*	-	9.52E-06	CAATTCTATAAGATTGA	0.78	0.73	0.25	0.55	0.52	0.25	0.55	0.64	0.27	0.40	0.45	0.93	0.40	0.55	0.52	0.14	0.25
	3	28864-28880	-69 *psbM*	-	1.30E-05	CATCTCTATGGGATTAA	0.78	0.73	0.25	0.23	0.52	0.25	0.55	0.64	0.27	0.28	0.17	0.93	0.40	0.55	0.52	0.70	0.25
	4	58225-58241	*accD* ORF	+	3.00E-05	CGGTTTTATGTGATTAT	0.78	0.22	0.25	0.55	0.52	0.25	0.55	0.64	0.27	0.28	0.18	0.93	0.40	0.55	0.52	0.70	0.25
	5	44932-44948	*trnS* 3'-end?	+	3.11E-05	CAATCTTACGTGATTGA	0.78	0.73	0.25	0.55	0.25	0.25	0.55	0.64	0.69	0.28	0.18	0.93	0.40	0.55	0.52	0.14	0.25
	6	10135-10151	*atpA* ORF	-	3.80E-05	CATTTATACCGGAACAA	0.78	0.73	0.25	0.55	0.52	0.25	0.55	0.64	0.69	0.19	0.17	0.93	0.40	0.15	0.25	0.70	0.25
	7	75581-75597	*petB intron*	+	4.33E-05	GAGCTGTACGAGATGAA	0.14	0.73	0.25	0.23	0.52	0.25	0.55	0.64	0.69	0.28	0.45	0.93	0.40	0.55	0.14	0.70	0.25
	8	46706-46722	non-coding region	-	4.93E-05	CAATTCAATATGGTTAT	0.78	0.73	0.25	0.55	0.52	0.25	0.15	0.64	0.27	0.40	0.18	0.93	0.28	0.55	0.52	0.70	0.25
	9	85895-85911	*rpl23* ORF	-	5.37E-05	CGCTTCAACCGGGTTAT	0.78	0.22	0.25	0.55	0.52	0.25	0.15	0.64	0.69	0.19	0.17	0.93	0.28	0.55	0.52	0.70	0.25
	10	152738-152754	*rpl23* ORF	+	5.37E-05	CGCTTCAACCGGGTTAT	0.78	0.22	0.25	0.55	0.52	0.25	0.15	0.64	0.69	0.19	0.17	0.93	0.28	0.55	0.52	0.70	0.25

C



D

EMB175	S	P	L	S	P	L	S	S	P	L	S	P	L	S	P	L2	S2
aa 6	N	T	V	N	N	S	N	T	N	T	T	T	T	N	N	T	T
aa 1'	T	N	S	D	D	E	D	N	T	N	S	D	D	D	D	T	E
rpl16 5' end sequence	C	A	A	T	T	C	T	A	T	A	A	G	A	T	T	G	A
Code matching																	

PPR103	S	P	L	S	P	L	S	S	P	L	S	P	L	S	P	L2	S2
aa 6	N	S	V	N	N	S	N	T	N	T	A	N	T	N	N	L	A
aa 1'	D	T	L	D	D	D	D	N	S	S	F	D	A	D	D	A	G
rpl16 5' end sequence	T	C	T	T	T	C	T	A	T	A	A	G	A	T	T	T	A
Code matching																	

Figure 3. PPR code-based prediction of EMB175/AtPPR103 binding sites. (**A**) Nucleotide binding probabilities for EMB175 PPR motifs (P, L and S) based on the amino acids found at the two primary specificity determining positions (amino acid 6 and 1′) (see Supplementary Figure S1). Repeats are listed from N to C-terminus. Probabilities are based on correlations between alignments between PLS editing factors and their inferred binding sites, as described in (19). (**B**) Prediction of EMB175 binding sites within the Arabidopsis chloroplast genome. The ten top ranking matches among both strands of the complete chloroplast genome are shown. The arrowhead marks the site in the *rps3-rpl16* intergenic region, shown in subsequent experiments to be an *in vivo* target of PPR103. The genomic location (NC_000932.1) and nucleotide sequence of each site are indicated, along with the binding score for each repeat. The *P*-values were calculated with the FIMO program (27). (**C**) Multiple sequence alignment of the *rps3-rpl16* intergenic region from Zea mays (Zm), Arabidopsis thaliana (At), Nicotiana tabacum (Nt) and Oryza sativa (Os). The putative EMB175 binding site upstream of *rpl16* and the sRNA representing a likely PPR footprint in maize (see Figure 5) are underlined with solid and dashed lines, respectively. (**D**) Alignment of the PPR motifs in EMB175 and PPR103 with their putative binding site upstream of *rpl16*. The two specificity determining amino acids (aa) in each PPR motif (see Supplementary Figure S1) are shown. Highly correlated matches are marked in black and weaker but significant matches are marked in gray.

maining PPR motifs did not reveal potential targets that could explain the ribosome defect in *ppr103* mutants. In addition, there is no apparent similarity between the predicted PPR103 binding site and the *cis*-element that is expected to specify editing at the *rpl2* start codon, supporting the view that the partial *rpl2* editing defect in *ppr103* mutants is a secondary effect (Supplementary Figure S2B). The PPR motifs in the Arabidopsis PPR103 ortholog, EMB175, show fewer irregularities (Figure 3D and Supplementary Figure S1). Therefore, we took the approach of predicting binding sites for EMB175 and then prioritized candidate sites for follow-up based on phylogenetic conservation with the orthologous sites in maize.

The predicted EMB175 binding site (Figure 3A) was used to query the complete Arabidopsis chloroplast genome. Matches with the lowest *P*-values are shown in Figure 3B. A match in the *rps3-rpl16* intergenic region stood out because (i) it had the second lowest *P*-value and it shows the longest contiguous set of matches to the predicted EMB175 binding site of any sequence in the chloroplast genome; (ii) it maps to an intergenic region in a polycistronic transcription unit, a common site of action for characterized PPR proteins in chloroplasts (3); (iii) the sequence of this region is well conserved among monocot and dicot species (Figure 3C), as is often true for PPR binding sites in chloroplasts (28,42,43). The other top matches mapped either to

the noncoding strand or their sequences were not conserved in maize.

PPR103 stabilizes processed dicistronic *rpl16-rpl14* mRNAs

The analyses above point to the sequence in the *rps3-rpl16* intergenic region as the best candidate for a direct binding site for PPR103. To test whether PPR103 influences the metabolism of RNA from this region, transcripts from this transcription unit were investigated in *ppr103* mutants by RNA gel blot hybridization (Figure 4). Because severe defects in plastid translation cause pleiotropic effects on RNA metabolism (44), we compared RNA from strong and weak *ppr103* alleles (*ppr103-2/-3* and *ppr103-1/-2,* respectively) to RNAs from two other mutants with plastid rRNA deficiencies of similar magnitude (*ppr4* and *hcf7,* respectively). Analysis of null mutants (*ppr103-2/-3)* with a probe for the *rpl16* exon showed the absence of two prominent transcripts at 1 and 2 kb, both of which accumulated normally in the *ppr4* mutant control (Figure 4A). Based on prior analyses of transcripts from this region (24), the affected transcripts were expected to be spliced and unspliced isoforms of a dicistronic *rpl16-rpl14* transcript. Analysis of RNA from the hypomorphic allele combination *ppr103-1/-2,* confirmed this to be the case (Figure 4B): the transcripts missing in *ppr103* mutants hybridize to probes for *rpl16* exon 2, the *rpl16* intron and *rpl14,* but not to transcripts from flanking genes. The loss of processed *rpl16-rpl14* RNAs in *ppr103* mutants was not accompanied by an increased level of most of the RNA precursors (compare *ppr103* mutants to the *hcf7* and *ppr4* controls), arguing that PPR103 stabilizes these RNAs rather than promoting their processing. However, it is possible that a transcript at ∼7 kb accumulates to increased levels in *ppr103* mutants so a defect in RNA cleavage cannot be completely ruled out.

PPR103 defines the 5′-end of processed *rpl16* mRNA

Processed RNA termini in chloroplasts are stabilized primarily by the site-specific binding of PPR (or PPR-like) proteins that block exoribonucleolytic degradation (16,28,43,45,46). To explore the possibility that PPR103 acts in this manner, we used a primer extension assay to map 5′ ends in the *rps3-rpl16* intergenic region, and to quantify their abundance in *ppr103* mutants (Figure 4C). One 5′ end was detected, which mapped 54 nucleotides upstream of the *rpl16* start codon. Transcripts with this end are strongly diminished in a hypomorphic *ppr103* mutant but accumulate normally in *hcf7* mutants, which have a plastid ribosome deficiency of similar magnitude. Results of a cRT-PCR assay confirmed this *rpl16* 5′-end to be the major one accumulating in maize chloroplasts, and also mapped the processed 3′-end downstream of *rpl14* (Supplementary Figure S3). The calculated size of spliced and unspliced *rpl16-rpl14* transcripts based on these mapped termini are 1017 and 2059 nucleotides, respectively, which match the sizes of the two major PPR103-dependent transcripts detected on northern blots (Figure 4).

Additional evidence that the sequence near *rpl16* may be bound by PPR103/EMB75 comes from an analysis of chloroplast small RNAs (sRNAs). The RNA segments bound by some PPR proteins accumulate *in vivo* as sRNAs, due to protection by the bound protein (16,27,40). This is best documented for proteins that stabilize processed mRNA termini, in which case the boundaries of the stabilized sRNAs correspond with the termini of the stabilized mRNA isoform(s). We detected an abundant sRNA in maize chloroplasts that spans the predicted PPR103 binding site and that has features of a PPR footprint (sharp 5′ boundary, conserved sequence, low secondary structure) (Figure 5A). The 5′ end of this sRNA matches that of the transcripts that require PPR103 for their accumulation. These results suggested that the sRNA constitutes PPR103's *in vivo* RNA footprint. To further address this possibility, we quantified this sRNA in weak (*ppr103-1*, *ppr103-1/-2*) and strong (*ppr103-2/-3*) *ppr103* alleles by RNA gel blot hybridization using an oligonucleotide probe complementary to the sRNA sequence (Figure 5B). RNA from several other mutants were included as controls. The *atp4* mutant is a particularly suitable control for this experiment because it lacks the same *rpl16-rpl14* mRNAs as *ppr103* but ATP4 is believed to promote the stabilization of the 3′ end of this dicistronic mRNA rather than its 5′ end (24). The *hcf7* and *ppr4* mutants exhibit ribosome deficiencies similar in magnitude to those in the weak and strong *ppr103* alleles, respectively. The results showed a reduction in the abundance of the sRNA in hypomorphic *ppr103* mutants and a complete loss of the sRNA in the strong *ppr103* mutant as compared to wild-type, *hcf7, ppr4* and *atp4* mutants. This observation together with other results presented above provides strong evidence that the sequence represented in this sRNA constitutes an *in vivo* binding site for PPR103, and that binding to this sequence in the context of unprocessed *rpl16* transcripts defines the position of the processed *rpl16* 5′ end while also stabilizing the downstream RNA.

Recombinant PPR103 binds with specificity to the 5′-end of processed *rpl16* mRNA

To confirm that the sRNA that maps to the 5′-end of *rpl16* is PPR103's RNA footprint, we generated recombinant PPR103 (rPPR103) fused to a maltose-binding protein (MBP) tag (Figure 6A) and measured the RNA binding activity of this protein with gel mobility shift assays. The affinity of the protein for an RNA corresponding to the PPR103-dependent sRNA was compared to that for two other RNAs of similar length (RNA1 and 2). The binding reactions included 0.5 mg/ml heparin to reduce nonspecific interactions. rPPR103 bound with much higher affinity to the RNA corresponding to the sequence of *rpl16* sRNA than to the unrelated RNAs (Figure 6B). Residual binding could be observed for RNA2 only at the highest rPPR103 concentration. No binding activity was detected with purified MBP at a concentration equivalent to the highest concentration of MBP-PPR103 used in the binding assays, demonstrating that it is the PPR103 moiety that harbors the RNA-binding activity. The binding specificity of rPPR103 was further explored by competition assays in which binding to the radiolabeled *rpl16* ligand was challenged by the addition of unlabeled RNA competitors (RNA1, *rpl16* and RNA2) (Figure 6C). The unlabeled *rpl16* RNA inhibited

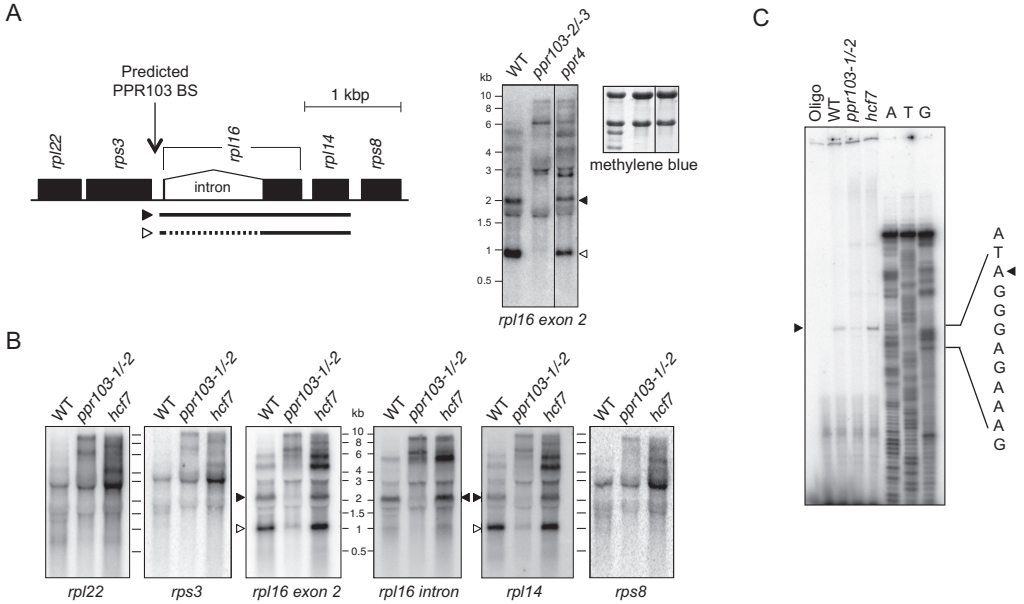


Figure 4. RNA gel blot analysis of *rpl16* RNAs in *ppr103* mutants and mapping of processed *rpl16* mRNA 5′-end. (**A**) RNA gel blot hybridization showing loss of specific *rpl16* transcripts in strong *ppr103* mutants. The diagram displays genes surrounding *rpl16* in angiosperm chloroplast genomes and the position of the predicted PPR103 binding site (BS). Seedling leaf RNA (5 μg) from a plant harboring the strong *ppr103-2/-3* allele was compared to that from a control mutant, *ppr4,* that also has a severe plastid ribosome deficiency. The blot was hybridized with an *rpl16* exon 2 probe. The lanes separated by a line come from non-adjacent lanes on the same exposure of the same blot. (**B**) RNA gel blots of seedling leaf RNA (5 μg) from plants harboring the weak *ppr103-1/-2* allele and a control mutant, *hcf7,* with a moderate loss of plastid ribosomes. The blots were hybridized with the indicated probes. Black and white arrowheads indicate the unspliced and spliced forms of processed *rpl16-rpl14* transcripts, respectively. (**C**) Primer extension analysis of the processed *rpl16* 5′-end in maize chloroplasts. The ddA, ddT and ddG sequencing ladders identify the positions of U, A and C residues in the RNA template. RNA samples from WT, *ppr103-1/-2* and *hcf7* were analyzed. An arrowhead indicates the major *rpl16* 5′-end.

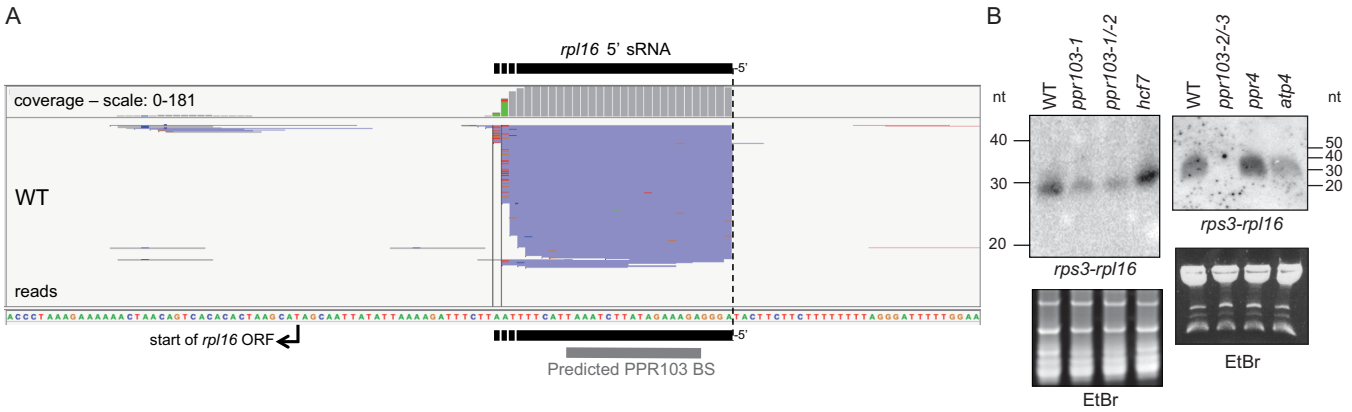


Figure 5. An sRNA corresponds to the 5′ end of PPR103-dependent *rpl16* transcripts. (**A**) RNA sequencing reads showing an abundant sRNA derived from sequences upstream of *rpl16* in the maize chloroplast genome. Sequencing reads from an sRNA library generated from Zea mays B73 leaf RNA were aligned to the maize chloroplast genome (NC_001666) using the IGV software. The *rps3-rpl16* sRNA sequence and the predicted PPR103 binding site (BS) are underlined in black and gray, respectively. (**B**) RNA gel blot demonstrating the accumulation of a PPR103-dependent sRNA. The accumulation of the sRNA shown in panel A was assessed in weak (*ppr103-1/-2*, *ppr103-1*) and strong (*ppr103-2/-3*) *ppr103* alleles along with control mutants *hcf7*, *ppr4* and *atp4.* The ethidium bromide (EtBr) stained gels are shown below to illustrate equal sample loading.

the binding of rPPR103 binding to labeled *rpl16* at a lower concentration than did RNAs 1 and 2 (Figure 6C). Our results demonstrated that rPPR103 binds with specificity to an RNA sequence that accumulates in a PPR103-dependent fashion *in vivo*. Taken together, the *in vivo* and *in vitro* data provide strong evidence that PPR103 binds to the *rps3-rpl16* intergenic region to define and stabilize the 5′-end of processed *rpl16* RNAs.

DISCUSSION

The results presented here demonstrate molecular and physiological functions for the PLS-type PPR protein PPR103. Although the vast majority of characterized PLS-PPR proteins specify sites of organellar RNA editing, PPR103 promotes the accumulation of dicistronic *rpl16-rpl14* transcripts. We provide strong evidence that this effect is mediated by the binding of PPR103 to sequences mapping be-

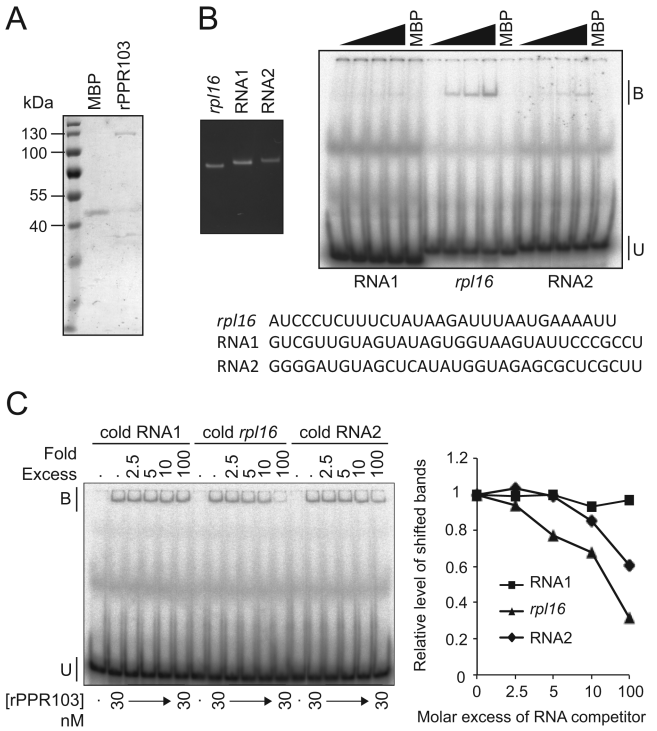


Figure 6. Gel mobility shift assays showing preferential RNA binding of recombinant PPR103 (rPPR103) to *rpl16* sRNA. (**A**) Purification of rPPR103. 100 and 200 ng of purified rPPR103 and MBP, respectively, were analyzed by SDS-PAGE and staining with Coomassie Brilliant Blue. The predicted sizes of rPPR103 and MBP are 129 and 44 kDa, respectively. (**B**) Gel mobility shift assays with rPPR103. The RNAs used in the binding assays were gel purified and ∼100 ng of each was resolved on a 12% denaturing polyacrylamide gel and stained with ethidium bromide to assess their purity (left panel), and 30 pM of radiolabeled RNAs were incubated with increasing concentrations of rPPR103 (0, 7.5, 15, 30 nM) or MBP (30 nM). The binding assays were run on a native acrylamide gel. Bound (B) and unbound (U) RNAs are indicated. The sequences of the RNAs are shown below. (**C**) Gel mobility shift assays using unlabeled RNA competitors. The rPPR103 concentration was kept constant (30 nM) and the molar excess of cold RNAs relative to the labeled *rpl16* RNA (30 pM) is indicated at the top of the gel. Quantification of the amount of radioactive RNA in the bound fraction is shown to the right.

tween ∼-54 and -27 with respect to the *rpl16* start codon. The most parsimonious interpretation of our results is that PPR103 bound to this site serves as a steric block to 5′→3′ exoribonucleolytic RNA decay, as shown previously for the PPR protein PPR10 (16,45), the HAT repeat protein HCF107 (46,47), and inferred for many other PPR and PPR-like proteins (reviewed in 3,28,43).

Blurring the line between the functional repertoire of P-type and PLS-type PPR proteins

We showed that PPR103, a PLS-PPR protein, protects and defines the 5′ end of processed *rpl16* transcripts. Although many other PPR proteins act analogously at other RNA termini, this type of activity has been attributed primarily to "pure" PPR proteins (P-PPR) that harbor long tracts of canonical PPR motifs (reviewed in 3). Those that have been characterized biochemically bind RNA with extremely high affinity and specificity (28,45,48), presumably due to their long contiguous RNA binding surface. The strength of these interactions is reflected by the fact that the "footprints" of such proteins accumulate *in vivo* as sRNAs due to protection by the protein from ribonuclease attack. The mining of plant sRNA data revealed many sRNAs with features of PPR footprints, some of which map to the genetically defined sites of action of P-PPR proteins (28,43). However, sites bound by the many PLS-PPR proteins known to be involved in RNA editing are not represented by sRNAs, implying lower affinity binding to RNA. This is consistent with their role in RNA editing, which generally occurs within open reading frames where high affinity interactions might inhibit translation. Our finding that a PLS-PPR protein has a molecular barrier activity similar to that of many P-PPR proteins challenges the view that the PLS repeat architecture is intrinsically less capable of achieving high affinity RNA interactions. This possibility was foreshadowed by genetic data for the PLS-PPR protein CRR2, which is involved in the intercistronic RNA stabilization/cleavage of *rps7/ndhB* transcripts in Arabidopsis (49); however the CRR2 binding site and mechanism are unknown. In any case, proteins with the PLS architecture should now be considered viable candidates for protecting the many other putative PPR footprints that have been cataloged in sRNA sequencing studies.

Comparison of the amino acid sequences found in the P, L and S PPR motifs of EMB175 and PPR103 in relation to their target RNA sequences provides insight into the role of each motif in RNA binding (Figure 3D). The amino acids found at the canonical specificity-determining positions (6 and 1′) in three of the four L motifs in PPR103/EMB175 do not correlate with the PPR code as established for P motifs (see Figure 3D). Prior reports provided evidence that L motifs in PLS-PPR RNA editing factors do not contribute to sequence-specific RNA recognition (6,50), and our results suggest the same is true for the L motifs in PPR103/EMB175. L motifs in EMB175/PPR103 might serve as spacers between S and P motifs to allow their correct alignment with the RNA bases they contact. Indeed, the amino acid combinations found in most of the S and P motifs in EMB175 and PPR103 align to the target RNA as predicted by the PPR code (Figure 3D). Biochemical and computational analysis of PLS-PPR RNA editing factors support the view that S and P motifs can participate in base recognition via a code that is the same as that for P motifs in "pure" PPR proteins (6,50,51). Interestingly, motifs 2, 11 and 12 match the code in EMB175 but not in PPR103. These motifs in PPR103 may bind bases in a noncanonical fashion; alternatively, these motifs may not make a strong contribution to RNA specificity/affinity even in EMB175. The latter possibility is consistent with prior reports that certain P and S motifs make little apparent contribution to RNA binding (50,52).

Altogether, our observations support the idea that the RNA stabilization factor PPR103 binds and recognizes its RNA target via a mechanism that is similar to that for PLS-PPR proteins involved in RNA editing (7,19).

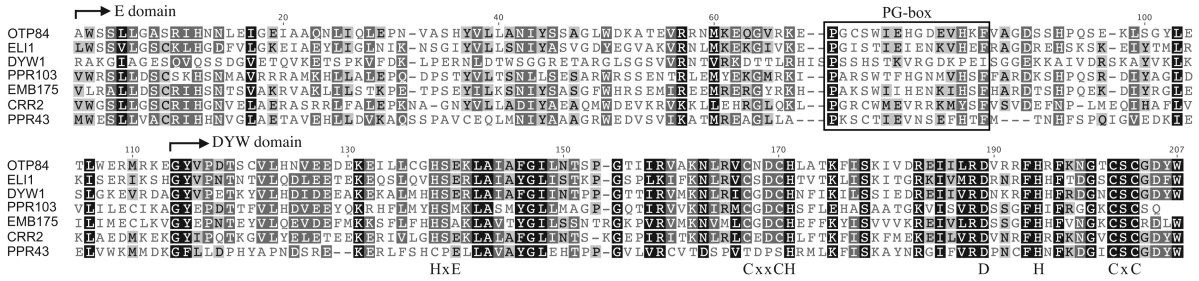


Figure 7. Multiple sequence alignment between E-DYW domains from PLS-PPR factors with distinct molecular functions. OTP84 (AT3G57430) (64), ELI1 (AT4G37380) (35) and DYW1 (AT1G47580) (65) are plastid editing factors in Arabidopsis, CRR2 (AT3G46790) functions in plastid RNA cleavage and/or stabilization in Arabidopsis (49) and PPR43 (Pp1s446_7V6) functions in mitochondrial RNA splicing in Physcomitrella patens (57). The PG-box that has been shown to be critical for plastid RNA editing and the residues in the DYW domain that are involved in zinc binding are marked (34,35).

PPR103 has an unusual DYW domain

PPR103 harbors a C-terminal extension, composed of an E and DYW domain (the latter so-named for its conserved C-terminal Asp-Tyr-Trp tripeptide) (Figure 1A and Supplementary Figure 1). Many PLS-PPR involved in organellar RNA editing carry a DYW domain and various lines of evidence suggest that it binds zinc and contributes to catalysis in the RNA editing reaction (33–36,53,54). However, the DYW domain is dispensable for the *in vivo* function of several PPR editing factors and *in vitro* assays of proteins harboring this domain have failed to detect any deaminase activity (55,56). PPR103 is the third DYW-domain containing PPR protein reported to function in a process other than RNA editing. CRR2 is involved in processing or stabilization of specific *rps7/ndhB* transcripts in Arabidopsis chloroplasts (49), and PpPPR_43 promotes splicing of *cox1* intron 3 in moss mitochondria (57). A comparison of the amino acid sequence of the E-DYW domain from CRR2, PpPPR_43 and PPR103/EMB175 with those in several editing factors revealed that both PpPPR_43 and PPR103/EMB175 proteins lack conserved residues that are expected to be key features for DYW "deaminases" (Figure 7). In particular, PPR103/EMB175 lack the conserved glutamate residue in the HxE motif that is proposed to facilitate the nucleophilic attack in the deamination process, as well as conserved residues in the PG-box found in chloroplastic editing factors (36). These sequence features strongly suggest that PPR103/EMB175 and PpPPR_43 evolved from editing factors, and that their acquisition of novel functions that do not require a catalytic center is linked with the evolutionary degeneration of their DYW domains.

The role of PPR103/EMB175 in plastid translation

The chloroplast *rpl16* and *rpl14* genes encode components of the chloroplast ribosome whose bacterial orthologs are essential for translation (58, reviewed in 59). Thus, the loss of dicistronic *rpl16-rpl14* transcripts in *ppr103* mutants appears at first glance to be a satisfactory explanation for the loss of plastid ribosomes in *ppr103* mutants. However, another maize PPR mutant, *atp4*, lacks exactly the same mRNA isoforms but shows no more than a minimal decrease in plastid ribosomes (24). Two explanations seem plausible for this apparent discrepancy: (i) PPR103 may have an additional, as yet undetected function that stimulates the expression of another chloroplast gene involved in translation, or (ii) PPR103 may promote the expression of *rpl16* and/or *rpl14* by doing more than simply stabilizing their dicistronic mRNA. In fact, several other PPR (and PPR-like) proteins that stabilize specifical processed 5′ termini also function as translational activators of the gene downstream from their binding site (3). Therefore, it is reasonable to consider the possibility that PPR103 activates *rpl16* translation whereas *atp4* (whose binding site is unknown) does not. In any case, the role of PPR103 in translation, if conserved in Arabidopsis, can account for the embryo-essential role of the Arabidopsis ortholog (EMB175), as plastid translation is essential for embryogenesis in Arabidopsis (60,61).

Utility of a code to predict PPR binding sites

The elucidation of an amino acid code that influences the nucleotide specificity of PPR motifs (6,7,19) offers the promise to design synthetic PPR proteins with desired RNA specificities (62,63), and to modulate the binding specificity of natural PPR proteins (6,50). In addition, the code constitutes a powerful tool for the prediction of RNA targets for the many uncharacterized PPR proteins in plants. Using the code, previous studies successfully predicted the RNA targets for PPR proteins involved in organellar RNA editing (7,19–21). In these studies, however, the PPR binding sites were predicted from a relatively limited sequence space: the *cis*-elements flanking mitochondrial and chloroplast editing sites. In this study, we used this code in conjunction with phylogenetic conservation to infer candidate-binding sites among the entire chloroplast genome. Molecular analyses of *ppr103* mutants confirmed that a physiologically relevant target of PPR103 was strongly predicted by this approach, and that PPR103 is required for the accumulation of RNAs harboring the predicted binding site at their 5′ end. This success adds to the evidence that the PPR recognition code can be used to accelerate the functional annotation of this large and essential gene family in plants.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We are grateful to Tiffany Kroeger for performing the reverse genetic screen that recovered the alleles described here, Susan Belcher for mutant propagation and complementation crossing, Roz Williams-Carrier for performing an sRNA northern blot, and Nicolas Baumberger and Laurence Herrgott for assistance with the FPLC.

FUNDING

CNRS; Marie Curie Integration Grant [CIG-618492-plantMTERF to K.H.]; US National Science Foundation [MCB-1243641 to A.B.]; Deutsche Forschungsgemeinschaft [TA 642/2-2, TA 642/6-1, TA 642/3-1 to M.T.]; Deutsche Forschungsgemeinschaft Heisenberg Fellowship [TA 642/4-1 to M.T.]. Funding for open access charge: Marie Curie Career Integration grant [CIG-618492].
Conflict of interest statement. None declared.

REFERENCES

1. Small,I.D. and Peeters,N. (2000) The PPR motif - a TPR-related motif prevalent in plant organellar proteins. *Trends Biochem. Sci*, **25**, 46–47.
2. Lurin,C., Andres,C., Aubourg,S., Bellaoui,M., Bitton,F., Bruyere,C., Caboche,M., Debast,C., Gualberto,J., Hoffmann,B. *et al.* (2004) Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell*, **16**, 2089–2103.
3. Barkan,A. and Small,I. (2014) Pentatricopeptide repeat proteins in plants. *Annu. Rev. Plant Biol.*, **65**, 415–442.
4. Yin,P., Li,Q., Yan,C., Liu,Y., Liu,J., Yu,F., Wang,Z., Long,J., He,J., Wang,H.W. *et al.* (2013) Structural basis for the modular recognition of single-stranded RNA by PPR proteins. *Nature*, **504**, 168–171.
5. Ban,T., Ke,J., Chen,R., Gu,X., Tan,M.H., Zhou,X.E., Kang,Y., Melcher,K., Zhu,J.K. and Xu,H.E. (2013) Structure of a PLS-class pentatricopeptide repeat protein provides insights into mechanism of RNA recognition. *J. Biol. Chem.*, **288**, 31540–31548.
6. Barkan,A., Rojas,M., Fujii,S., Yap,A., Chong,Y.S., Bond,C.S. and Small,I. (2012) A combinatorial amino acid code for RNA recognition by pentatricopeptide repeat proteins. *PLoS Genet.*, **8**, e1002910.
7. Yagi,Y., Hayashi,S., Kobayashi,K., Hirayama,T. and Nakamura,T. (2013) Elucidation of the RNA recognition code for pentatricopeptide repeat proteins involved in organelle RNA editing in plants. *PLoS One*, **8**, e57286.
8. Cushing,D.A., Forsthoefel,N.R., Gestaut,D.R. and Vernon,D.M. (2005) Arabidopsis emb175 and other ppr knockout mutants reveal essential roles for pentatricopeptide repeat (PPR) proteins in plant embryogenesis. *Planta*, **221**, 424–436.
9. Barkan,A., Walker,M., Nolasco,M. and Johnson,D. (1994) A nuclear mutation in maize blocks the processing and translation of several chloroplast mRNAs and provides evidence for the differential translation of alternative mRNA forms. *EMBO J.*, **13**, 3170–3181.
10. Meierhoff,K., Felder,S., Nakamura,T., Bechtold,N. and Schuster,G. (2003) HCF152, an Arabidopsis RNA binding pentatricopeptide repeat protein involved in the processing of chloroplast psbB-psbT-psbH-petB-petD RNAs. *Plant Cell*, **15**, 1480–1495.
11. Yamazaki,H., Tasaka,M. and Shikanai,T. (2004) PPR motifs of the nucleus-encoded factor, PGR3, function in the selective and distinct steps of chloroplast gene expression in Arabidopsis. *Plant J.*, **38**, 152–163.
12. Kotera,E., Tasaka,M. and Shikanai,T. (2005) A pentatricopeptide repeat protein is essential for RNA editing in chloroplasts. *Nature*, **433**, 326–330.
13. Loiselay,C., Gumpel,N.J., Girard-Bascou,J., Watson,A.T., Purton,S., Wollman,F.A. and Choquet,Y. (2008) Molecular identification and function of cis- and trans-acting determinants for petA transcript stability in Chlamydomonas reinhardtii chloroplasts. *Mol. Cell Biol.*, **28**, 5529–5542.
14. Johnson,X., Wostrikoff,K., Finazzi,G., Kuras,R., Schwarz Bujaldon,C. S., Nickelsen,J., Stern,D.B., Wollman,F.A. and Vallon,O. (2010) MRL1, a conserved Pentatricopeptide repeat protein, is required for stabilization of rbcL mRNA in Chlamydomonas and Arabidopsis. *Plant Cell*, **22**, 234–248.
15. Schmitz-Linneweber,C., Williams-Carrier,R. and Barkan,A. (2005) RNA immunoprecipitation and microarray analysis show a chloroplast Pentatricopeptide repeat protein to be associated with the 5′ region of mRNAs whose translation it activates. *Plant Cell*, **17**, 2791–2804.
16. Pfalz,J., Bayraktar,O.A., Prikryl,J. and Barkan,A. (2009) Site-specific binding of a PPR protein defines and stabilizes 5′ and 3′ mRNA termini in chloroplasts. *EMBO J.*, **28**, 2042–2052.
17. Schmitz-Linneweber,C., Williams-Carrier,R.E., Williams-Voelker,P.M., Kroeger,T.S., Vichas,A. and Barkan,A. (2006) A pentatricopeptide repeat protein facilitates the trans-splicing of the maize chloroplast rps12 pre-mRNA. *Plant Cell*, **18**, 2650–2663.
18. Beick,S., Schmitz-Linneweber,C., Williams-Carrier,R., Jensen,B. and Barkan,A. (2008) The pentatricopeptide repeat protein PPR5 stabilizes a specific tRNA precursor in maize chloroplasts. *Mol. Cell Biol.*, **28**, 5337–5347.
19. Takenaka,M., Zehrmann,A., Brennicke,A. and Graichen,K. (2013) Improved computational target site prediction for pentatricopeptide repeat RNA editing factors. *PLoS One*, **8**, e65343.
20. Yagi,Y., Tachikawa,M., Noguchi,H., Satoh,S., Obokata,J. and Nakamura,T. (2013) Pentatricopeptide repeat proteins involved in plant organellar RNA editing. *RNA Biol.*, **10**, 1419–1425.
21. Yap,A., Kindgren,P., Colas des Francs-Small,C., Kazama,T., Tanz,S.K., Toriyama,K. and Small,I. (2015) AEF1/MPR25 is implicated in RNA editing of plastid atpF and mitochondrial nad5, and also promotes atpF splicing in Arabidopsis and rice. *Plant J.*, **81**, 661–669.
22. Belcher,S., Williams-Carrier,R., Stiffler,N. and Barkan,A. (2015) Large-scale genetic analysis of chloroplast biogenesis in maize. *Biochim. Biophys. Acta.*, **1847**, 1004–1016.
23. Barkan,A. (1993) Nuclear mutants of maize with defects in chloroplast polysome assembly have altered chloroplast RNA metabolism. *Plant Cell*, **5**, 389–402.
24. Zoschke,R., Watkins,K.P. and Barkan,A. (2013) A rapid ribosome profiling method elucidates chloroplast ribosome behavior in vivo. *Plant Cell*, **25**, 2265–2275.
25. Barkan,A. (1998) Approaches to investigating nuclear genes that function in chloroplast biogenesis in land plants. *Method Enzymol.*, **297**, 38–57.
26. Roy,L.M. and Barkan,A. (1998) A SecY homologue is required for the elaboration of the chloroplast thylakoid membrane and for normal chloroplast gene expression. *J. Cell Biol.*, **141**, 385–395.
27. Grant,C.E., Bailey,T.L. and Noble,W.S. (2011) FIMO: scanning for occurrences of a given motif. *Bioinformatics*, **27**, 1017–1018.
28. Zhelyazkova,P., Hammani,K., Rojas,M., Voelker,R., Vargas-Suarez,M., Borner,T. and Barkan,A. (2012) Protein-mediated protection as the predominant mechanism for defining processed mRNA termini in land plant chloroplasts. *Nucleic Acids Res.*, **40**, 3092–3105.
29. Asakura,Y. and Barkan,A. (2006) Arabidopsis orthologs of maize chloroplast splicing factors promote splicing of orthologous and species-specific group II introns. *Plant Physiol.*, **142**, 1656–1663.
30. Busso,D., Delagoutte-Busso,B. and Moras,D. (2005) Construction of a set Gateway-based destination vectors for high-throughput cloning and expression screening in Escherichia coli. *Anal. Biochem.*, **343**, 313–321.
31. Tomcal,M., Stiffler,N. and Barkan,A. (2013) POGs2: a web portal to facilitate cross-species inferences about protein architecture and function in plants. *PLoS One*, **8**, e82569.
32. Shikanai,T. (2015) RNA editing in plants: machinery and flexibility of site recognition. *Biochim. Biophys. Acta.*, **1847**, 779–785.
33. Wagoner,J.A., Sun,T., Lin,L. and Hanson,M.R. (2015) Cytidine deaminase motifs within the DYW domain of two pentatricopeptide repeat-containing proteins are required for site-specific chloroplast RNA editing. *J. Biol. Chem.*, **290**, 2957–2968.
34. Boussardon,C., Avon,A., Kindgren,P., Bond,C.S., Challenor,M., Lurin,C. and Small,I. (2014) The cytidine deaminase signature HxE(x)n CxxC of DYW1 binds zinc and is necessary for RNA editing of ndhD-1. *New Phytol.*, **203**, 1090–1095.

35. Hayes,M.L., Giang,K., Berhane,B. and Mulligan,R.M. (2013) Identification of two pentatricopeptide repeat genes required for RNA editing and zinc binding by C-terminal cytidine deaminase-like domains. *J. Biol. Chem.*, **288**, 36519–36529.
36. Hayes,M.L., Dang,K.N., Diaz,M.F. and Mulligan,R.M. (2015) A conserved glutamate residue in the C-terminal deaminase domain of pentatricopeptide repeat proteins is required for RNA editing activity. *J. Biol. Chem.*, **290**, 10136–10142.
37. Majeran,W., Friso,G., Asakura,Y., Qu,X., Huang,M., Ponnala,L., Watkins,K.P., Barkan,A. and van Wijk,K.J. (2012) Nucleoid-enriched proteomes in developing plastids and chloroplasts from maize leaves: a new conceptual framework for nucleoid functions. *Plant Physiol.*, **158**, 156–189.
38. Maier,R.M., Neckermann,K., Igloi,G.L. and Kossel,H. (1995) Complete sequence of the maize chloroplast genome: gene content, hotspots of divergence and fine tuning of genetic information by transcript editing. *J. Mol. Biol.*, **251**, 614–628.
39. Peeters,N.M. and Hanson,M.R. (2002) Transcript abundance supercedes editing efficiency as a factor in developmental variation of chloroplast gene expression. *RNA*, **8**, 497–511.
40. Halter,C.P., Peeters,N.M. and Hanson,M.R. (2004) RNA editing in ribosome-less plastids of iojap maize. *Curr. Genet.*, **45**, 331–337.
41. Jenkins,B.D., Kulhanek,D.J. and Barkan,A. (1997) Nuclear mutations that block group II RNA splicing in maize chloroplasts reveal several intron classes with distinct requirements for splicing factors. *Plant Cell*, **9**, 283–296.
42. Hayes,M.L. and Mulligan,R.M. (2011) Pentatricopeptide repeat proteins constrain genome evolution in chloroplasts. *Mol. Biol. Evol.*, **28**, 2029–2039.
43. Ruwe,H. and Schmitz-Linneweber,C. (2012) Short non-coding RNA fragments accumulating in chloroplasts: footprints of RNA binding proteins? *Nucleic Acids Res.*, **40**, 3106–3116.
44. Williams,P.M. and Barkan,A. (2003) A chloroplast-localized PPR protein required for plastid ribosome accumulation. *Plant J.*, **36**, 675–686.
45. Prikryl,J., Rojas,M., Schuster,G. and Barkan,A. (2011) Mechanism of RNA stabilization and translational activation by a pentatricopeptide repeat protein. *Proc. Natl. Acad. Sci. USA*, **108**, 415–420.
46. Hammani,K., Cook,W.B. and Barkan,A. (2012) RNA binding and RNA remodeling activities of the half-a-tetratricopeptide (HAT) protein HCF107 underlie its effects on gene expression. *Proc. Natl. Acad. Sci. USA*, **109**, 5651–5656.
47. Felder,S., Meierhoff,K., Sane,A.P., Meurer,J., Driemel,C., Plucken,H., Klaff,P., Stein,B., Bechtold,N. and Westhoff,P. (2001) The nucleus-encoded HCF107 gene of Arabidopsis provides a link between intercistronic RNA processing and the accumulation of translation-competent psbH transcripts in chloroplasts. *Plant Cell*, **13**, 2127–2141.
48. Cai,W., Okuda,K., Peng,L. and Shikanai,T. (2011) PROTON GRADIENT REGULATION 3 recognizes multiple targets with limited similarity and mediates translation and RNA stabilization in plastids. *Plant J.*, **67**, 318–327.
49. Hashimoto,M., Endo,T., Peltier,G., Tasaka,M. and Shikanai,T. (2003) A nucleus-encoded factor, CRR2, is essential for the expression of chloroplast ndhB in Arabidopsis. *Plant J.*, **36**, 541–549.
50. Kindgren,P., Yap,A., Bond,C.S. and Small,I. (2015) Predictable alteration of sequence recognition by RNA editing factors from Arabidopsis. *Plant Cell*, **27**, 403–416.
51. Okuda,K., Shoki,H., Arai,M., Shikanai,T., Small,I. and Nakamura,T. (2014) Quantitative analysis of motifs contributing to the interaction between PLS-subfamily members and their target RNA sequences in plastid RNA editing. *Plant J.*, **80**, 870–882.
52. Fujii,S., Sato,N. and Shikanai,T. (2013) Mutagenesis of individual pentatricopeptide repeat motifs affects RNA binding activity and reveals functional partitioning of Arabidopsis PROTON gradient regulation3. *Plant Cell*, **25**, 3079–3088.
53. Salone,V., Rudinger,M., Polsakiewicz,M., Hoffmann,B., Groth-Malonek,M., Szurek,B., Small,I., Knoop,V. and Lurin,C. (2007) A hypothesis on the identification of the editing enzyme in plant organelles. *FEBS Lett.*, **581**, 4132–4138.
54. Iyer,L.M., Zhang,D., Rogozin,I.B. and Aravind,L. (2011) Evolution of the deaminase fold and multiple origins of eukaryotic editing and mutagenic nucleic acid deaminases from bacterial toxin systems. *Nucleic Acids Res.*, **39**, 9473–9497.
55. Okuda,K., Chateigner-Boutin,A.L., Nakamura,T., Delannoy,E., Sugita,M., Myouga,F., Motohashi,R., Shinozaki,K., Small,I. and Shikanai,T. (2009) Pentatricopeptide repeat proteins with the DYW motif have distinct molecular functions in RNA editing and RNA cleavage in Arabidopsis chloroplasts. *Plant Cell*, **21**, 146–156.
56. Nakamura,T. and Sugita,M. (2008) A conserved DYW domain of the pentatricopeptide repeat protein possesses a novel endoribonuclease activity. *FEBS Lett.*, **582**, 4163–4168.
57. Ichinose,M., Tasaki,E., Sugita,C. and Sugita,M. (2012) A PPR-DYW protein is required for splicing of a group II intron of cox1 pre-mRNA in Physcomitrella patens. *Plant J.*, **70**, 271–278.
58. Shoji,S., Dambacher,C.M., Shajani,Z., Williamson,J.R. and Schultz,P.G. (2011) Systematic chromosomal deletion of bacterial ribosomal protein genes. *J. Mol. Biol.*, **413**, 751–761.
59. Tiller,N. and Bock,R. (2014) The translational apparatus of plastids and its role in plant development. *Mol. Plant*, **7**, 1105–1120.
60. Meinke,D., Muralla,R., Sweeney,C. and Dickerman,A. (2008) Identifying essential genes in Arabidopsis thaliana. *Trends Plant Sci.*, **13**, 483–491.
61. Bryant,N., Lloyd,J., Sweeney,C., Myouga,F. and Meinke,D. (2011) Identification of nuclear genes encoding chloroplast-localized proteins required for embryo development in Arabidopsis. *Plant Physiol.*, **155**, 1678–1689.
62. Shen,C., Wang,X., Liu,Y., Li,Q., Yang,Z., Yan,N., Zou,T. and Yin,P. (2015) Specific RNA recognition by designer pentatricopeptide repeat protein. *Mol. Plant*, **8**, 667–670.
63. Coquille,S., Filipovska,A., Chia,T., Rajappa,L., Lingford,J.P., Razif,M.F., Thore,S. and Rackham,O. (2014) An artificial PPR scaffold for programmable RNA recognition. *Nat. Commun.*, **5**, 5729.
64. Hammani,K., Okuda,K., Tanz,S.K., Chateigner-Boutin,A.L., Shikanai,T. and Small,I. (2009) A study of new Arabidopsis chloroplast RNA editing mutants reveals general features of editing factors and their target sites. *Plant Cell*, **21**, 3686–3699.
65. Boussardon,C., Salone,V., Avon,A., Berthome,R., Hammani,K., Okuda,K., Shikanai,T., Small,I. and Lurin,C. (2012) Two interacting proteins are necessary for the editing of the NdhD-1 site in Arabidopsis plastids. *Plant Cell*, **24**, 3684–3694.